DOCUMENT RESUME

ED 411 247 TM 027 121

AUTHOR Jacobs, Victoria R.
TITLE Children's Understanding of Sampling in Surveys.
PUB DATE 1997-03-00
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Affective Behavior; Children; *Comprehension; *Elementary School Students; Evaluation Methods; Intermediate Grades; *Sampling; *Surveys; Tables (Data); Test Interpretation; *Thinking Skills

ABSTRACT

Two studies investigated upper elementary school students' informal understanding of sampling issues in the context of interpreting and evaluating survey results. The specific focus was on the children's evaluation of sampling methods and means of drawing conclusions from multiple surveys. In Study 1, 17 children were individually interviewed to categorize children's conceptions. In Study 2, 110 children completed paper-and-pencil tasks to confirm the response categories identified in Study 1 and to determine the prevalence of the response categories in a larger sample. Children evaluated sampling methods focusing on potential for bias, fairness, practical issues, or results. All children used multiple types of evaluation rationales, and the focus of their evaluations varied somewhat by context and type of sampling method (restricted, self-selected, or random). Children used affective (fairness) rationales more often in school contexts and rationales focused on results more often in out-of-school contexts. Children had more difficulty detecting bias with self-selected sampling methods than with restricted sampling methods because self-selection was initially the most fair (i.e., everyone had a chance to participate). Children preferred stratified random sampling to simple random sampling because they wanted to ensure that all types of individuals were included. When drawing conclusions from multiple surveys, children: (1) considered survey quality; (2) aggregated all surveys regardless of quality; (3) used their own opinions and ignored all survey data; or (4) refused to draw conclusions. Even when children were able to identify potential bias, they often ignored survey quality when drawing conclusions from multiple surveys. (Contains 5 tables and 45 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 247

Children's Understanding of Sampling in Surveys

Victoria R. Jacobs

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Victoria Jacobs

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM027121

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Chicago, IL
March 1997

BEST COPY AVAILABLE

Children's Understanding of Sampling in Surveys

Victoria R. Jacobs
California State University San Marcos

American Educational Research Association Annual Meeting,
March 1997
Chicago, IL

Two studies investigated upper elementary children's informal understanding of sampling issues in the context of interpreting and evaluating survey results. Specifically, these studies investigated children's evaluation of sampling methods and means of drawing conclusions from multiple surveys. In Study 1, 17 children were individually interviewed to categorize children's conceptions. In Study 2, 110 children completed paper-and-pencil tasks (a) to confirm the response categories identified in Study 1 and (b) to determine the prevalence of the response categories in a larger sample. Children evaluated sampling methods by focusing on (a) potential for bias, (b) fairness, (c) practical issues, or (d) results. All children used multiple types of evaluation rationales, and the focus of their evaluations varied somewhat by context and type of sampling method (restricted, self-selected, or random). Children used affective (fairness) rationales more often in school contexts and rationales focused on results more often in out-of-school contexts. Children had more difficulty detecting bias with self-selected sampling methods than restricted sampling methods because self-selection was initially the most fair (i.e., everyone had a chance to participate). Children preferred stratified random sampling to simple random sampling because they wanted to ensure that all types of individuals were included. When drawing conclusions from multiple surveys, children (a) considered survey quality, (b) aggregated all surveys regardless of quality, (c) used their own opinions and ignored all survey data, or (d) refused to draw conclusions. Even when children were able to identify potential bias, they often ignored survey quality when drawing conclusions from multiple surveys.

The problem of statistical illiteracy has been underscored by the increasing prevalence of statistics in everyday life. In recent years, there has also been an increased interest in educating the lay public about the uses and misuses of statistics in everyday situations (see, for example, Crossen, 1994; Dewdney, 1993; Paulos, 1995). In 1989, the National Council of Teachers of Mathematics recognized the necessity of statistical understanding by recommending the introduction of statistics instruction for grades K-12 in their *Curriculum and Evaluation Standards for School Mathematics*. The current project was designed to inform instructional efforts to help children develop statistical understanding.

This project explored children's informal understanding based on the theoretical perspective that instruction should start from children's understanding and then build on their thinking (Confrey, 1990; Hiebert & Carpenter, 1992). Precedence for this approach can be found in the research on out-of-school learning (Lave, 1988; Nunes, Schliemann, & Carraher, 1993; Resnick, 1987; Saxe, 1988; Scribner, 1984) and in several educational programs which have been able to build successful instructional programs based on an understanding of and respect for children's informal knowledge (see, for example, Carpenter & Fennema, 1992; Cobb et al., 1991; Empson, 1995; Hunt & Minstrell, 1994; Mack, 1990; Papert, 1980).

Most of the recent statistics education projects have concentrated exclusively on the skills and knowledge required to collect, organize, and describe data. The underlying assumption is that children who can collect, organize, and describe their own data effectively can interpret and evaluate others'

TM027121

statistics. There is, however, no evidence that these skills are automatically transferable. Furthermore, "while not all students of elementary statistics will become statisticians, virtually all students will use and apply statistics in their future professional or private lives" (Jaffe & Spirer, 1987, p. 4).

An understanding of statistical inference is essential for individuals to deal effectively with the statistics they encounter (e.g., in the media). Statistical inference forces the "intuitive scientist to take a step beyond characterizing individual observations or data samples. The step is using the data at hand to make inferences about the general characteristics or 'parameters' of the population from which those data were drawn" (Nisbett & Ross, 1980, p. 76). However, statistical inference is almost by definition imperfect -- all sampling introduces some error. Consequently, individuals need to understand potential threats to valid statistical inference.

One of the main determinants of the validity of statistical inference is *sampling*. Sampling can be done well or poorly. Good samples are *representative* of their populations while poor samples are *biased* or *unrepresentative*. Nisbett and Ross (1980) define *sampling bias* as "any sampling procedure that fails to yield values identical on the average to those produced by a random procedure" (p. 83). The danger of unrepresentative sampling is that the inferences to the population are invalid. This study investigated what children understand informally about sampling methods that lead to unrepresentative samples (and invalid statistical inference) versus those that lead to representative samples (and valid statistical inference). Specifically, a combination of paper-and-pencil tasks and interviewing was used to investigate upper elementary children's informal understanding of statistical sampling in the context of interpreting and evaluating survey results. While the NCTM Standards have inspired a flourish of research activity investigating how to best teach statistics in the elementary school, this research has focused almost exclusively on descriptive statistics and on helping children effectively collect, organize, and describe data (Day, Webb, Nabate, & Romberg, 1987; de Lange, Burrill, Romberg, & van Reeuwijk, 1993; de Lange & Verhage, 1992; Gal, Rothschild, & Wagner, 1989; Hancock, Kaput, & Goldsmith, 1992; Lajoie, Lavigne, & Lawless, 1993; Lehrer & Romberg, 1996; National Center for Research in Mathematical Sciences Education & Freudenthal Institute, in press; Russell & Friel, 1989; TERC, in press). In contrast, little research has examined the areas of inferential statistics and helping children learn to interpret and evaluate statistics that others have created. This project addressed both of these under-researched topics.

While surveys are only one of many types of statistical information, they were chosen as the focus of this project because of their prevalence in the media and their accessibility to children. For example, in this project, children in three upper elementary classes were asked to find surveys *anywhere*. Almost every child said that it took only 5-10 minutes to locate a survey. Furthermore, the surveys they found were from a wide variety of places including the mail, newspapers, magazines, television shows, telephone inquiries, and cereal boxes.

Statistical Understanding in the General Population

The research on reasoning under uncertainty has shown that adults are generally poor statisticians (Kahneman, Slovic & Tversky, 1982). In particular, Nisbett and Ross (1980) found that adults are

lacking in their understanding of sampling issues and are often unable to identify samples that are unrepresentative because of inappropriate sampling methods. They suggest that this insensitivity to sampling methods may be due to individuals' misconceptions of the statistical meaning of *random*[1] for which they equate random with *haphazard* or *without pattern* . Nisbett and Ross (1980) warn that an inability "to understand the inferential advantages of a random selection procedure automatically betokens an insensitivity to the *disadvantages* of a *biased* sampling procedure" (p. 83).

In contrast to the large body of research on adult statistical (mis)conceptions, researchers and educators have little knowledge about children's conceptions of sampling issues. However, the little research that does exist has found that children also lack complete understanding (Lajoie, Jacobs, & Lavigne, 1995; Shaughnessy, 1992).

Influence of Non-Statistical Issues on Statistical Understanding

Little is known about how non-statistical issues such as context specifically affect statistical understanding. Nonetheless, researchers have found that children's understanding of probability is more limited in real-world contexts than in the contexts of standard probability tasks such as dice and coin flips (Garfield, 1995; Jacobs, 1993; Schwartz et al., 1994). Given these discrepancies and the societal goal of having children understand real-world statistics, this project focused on children's understanding in real-world contexts.

The use of real-world contexts increases the likelihood that individuals will have prior knowledge and beliefs about the issues being discussed. The research on prior knowledge and beliefs suggests that individuals might not be equally critical of all sampling methods. Specifically, they may be less critical of sampling methods that resulted in conclusions consistent with their prior knowledge and beliefs, and more critical of those that are not consistent (Lord, Ross, & Lepper, 1979). This asymmetrical interpretation and evaluation of evidence has been investigated under the name of *biased assimilation.*

Overview of Studies

Two studies investigated upper elementary children's informal understanding of statistical sampling issues in the context of interpreting and evaluating survey results. Specifically, these studies were designed to address the following questions:

1. How do children evaluate different sampling methods?
2. How do children draw conclusions from multiple surveys with conflicting results on the same topic?

The second question was included to determine if children consider the validity of each survey's data (i.e., survey quality) when drawing conclusions from multiple surveys. The particular conclusions drawn were of less interest than how children arrived at the conclusions. The methodologies of these two studies compliment each other by addressing the same issues through different lenses. Study 1 provided in-depth information made possible by the use of interviews to identify categories of children's

---

[1] The statistical meaning of simple random sampling refers to a sampling procedure in which "each member of the population has an equal chance of being included in the sample" and "each member is selected independently of all other members" (Marascuilo & Serlin, 1988, p. 14).

conceptions. Study 2 used paper-and-pencil tasks to provide a wide-angle lens that examined a larger sample of children's thinking. It was designed (a) to confirm the response categories identified in Study 1 and (b) to determine the prevalence of the response categories in a larger sample of children.

Both studies investigated children's reasoning behind their evaluations. This approach is in contrast to the dominant methodology in this field which uses forced-choice questions to assess what decisions individuals make in uncertain situations. Shaughnessy (1992) criticized the forced-choice approach for over-emphasizing "correct answers" and not probing participants about how they arrived at their solutions.

STUDY 1

Study 1 was designed to identify categories of children's conceptions of sampling from interviews. It was *not* intended to categorize children but rather to discover the range of ideas that children provided when discussing sampling issues in the context of interpreting and evaluating surveys.

Selection of Participants

Study 1 took place in a medium-sized city in Wisconsin in three classes which were selected due to the teachers' emphasis on having children explain their thinking. Study 1 used both paper-and-pencil and interview tasks that were developed for this study. All tasks were open-ended in nature so it was important to have children who were used to being questioned about their thinking. Teachers reported that the children had already been exposed to surveys through the class newspaper, but these surveys had not involved sampling (i.e., they always asked everyone in the class). Furthermore, teachers reported that they had not discussed any issues related to sampling prior to the study.

Paper-and-pencil screening. Thirty-one fourth graders (17 boys and 14 girls) and 32 fifth graders (14 boys and 18 girls) from three multi-age (grades 4 and 5) classrooms completed paper-and-pencil tasks during the month of April so that their responses could be used to select an appropriate subset of children to interview. In other words, the paper-and-pencil tasks were used as a screening task to ensure that the children interviewed demonstrated a range of understanding. The tasks were open-ended and asked children to describe their experiences with surveys, define survey terminology (e.g., survey, sample, random), and evaluate surveys presented in multiple contexts.

The screening tasks were used to categorize children in order to select children to interview. Analysis showed that children fell roughly into four categories:

1. Children who consistently focused on sampling factors that would affect the quality of surveys (e.g., sampling method, sample size).
2. Children who consistently focused on their own opinions when addressing the quality of surveys.
3. Children who were inconsistent in their focus. Sometimes they focused on sampling factors and other times they focused on their own opinions.
4. Children who focused on practical issues or were unsure about the tasks or how to address the quality of surveys.

Seven children were selected from category 1, six from category 3, and two from categories 2 and 4. Although it was important to represent children who showed limited understanding (categories 2 and 4), more children were selected from the other categories to better define the different facets of understanding when it did exist.

Interviews. Eight fourth graders (three boys and five girls) and nine fifth graders (five boys and four girls) were selected to interview during the month of May. The final selection of interviewees was done in conjunction with the children's classroom teachers. The goal was to select children with a range of understanding who would not be intimidated by the interviewing situation and would be verbal enough to describe their understanding. Care was also taken to ensure that both genders, all three classes, and a range of mathematical abilities were represented.

Interview Tasks

A variety of questions were used to assess children's understanding of sampling issues. All questions were deeply contextualized in either a school context or an out-of-school context with a population larger than a single school. For each context, there were several scenarios that each described a set of circumstances and a decision that needed to be made based on survey data. All scenarios described each survey by identifying the sampling method, sample size, and survey results. Children were asked to evaluate different sampling methods and/or draw conclusions from multiple surveys.

Three basic sampling methods were included: random, restricted, or self-selected sampling methods. ***Random sampling methods*** gave each member of the population the same chance of being selected. ***Restricted sampling methods*** asked particular groups of people who might be more likely to select a certain response and consequently, skew the results in a particular direction. ***Self-selected sampling methods*** had the participants select themselves and were problematic because there was no means of evaluating whether the sample was representative. Furthermore, individuals who choose to participate in a survey are likely to have different opinions than those who do not choose to participate (Asher, 1988).

Children were asked to evaluate sampling methods and draw conclusions in several scenarios. The following sections explain two of the major scenarios: raffle and recycling scenarios.

Raffle scenario. The raffle survey was a school context in which both sampling methods and results varied while the sample size remained constant. Each child being interviewed was provided with information similar to the following:

> The school is an elementary school with grades 1 through 6 and 100 students in each grade. A fifth grade class is trying to raise some money to go on a field trip to Great America (an amusement park). They are considering several options to raise money and decide to do a survey to help them determine the best way to raise the most money. One option is to sell raffle tickets for a SEGA™ video-game system. Consequently, nine different students each conducted a survey to estimate how many students in the school would buy a raffle ticket to win a SEGA™. Each survey asked 60 students but each sampling method and results were different. The nine surveys and their results were as follows:

1. Raffi asked 60 friends. (75% yes, 25% no)

2. Shannon got the names of all 600 kids in the school, put them in a hat, and pulled out 60 of them. (35% yes, 65% no)

3 Spence had blond hair so he asked the first 60 kids he found who had blond hair too. (55% yes, 45% no)

4. Jake asked 60 kids at an after school meeting of the Games Club. The Games Club met once a week and played different games -- especially computerized ones. Anyone who was interested in games could join. (90% yes, 10% no)

5. Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her. (50% yes, 50% no)

6. Claire set up a booth outside the lunchroom and anyone who wanted to could stop by and fill out her survey. To advertise her survey she had a sign that said "WIN A SEGA". She stopped collecting surveys when she got 60 completed. (100% yes)

7. Brooke asked the first 60 kids she found whose telephone number ended in a 3 because 3 was her favorite number. (25% yes, 75% no)

8. Kyle wanted the same number of boys and girls and some kids from each grade. So he asked 5 boys and 5 girls from each grade to get his total of 60 kids. (30% yes, 70% no)

9. Courtney didn't know too many boys so she decided to ask 60 girls. But she wanted to make sure she got some young girls and some older ones so she asked 10 girls from each grade. (10% yes, 90% no)

The child being interviewed was shown a description of the sampling method, sample size, and results for each of the nine surveys. Each child was then asked to evaluate these different surveys with open-ended questions such as "What do you think about [one of the nine surveys]?"; "If you could pick any way to select 60 kids, what would you do?"; or "What do you think is the best estimate of what percentage of kids will buy a raffle ticket?"

Recycling scenario. The recycling scenario was an out-of-school context in which sampling methods, sample sizes, and results all varied. Each child being interviewed was provided with information similar to the following:

> Two surveys were conducted to determine how many of the schools in Wisconsin are recycling. The first survey used a large sample size and a self-selected sampling method by sending out postcards to all the school principals in Wisconsin. About half of the principals sent them back, and 91% of those that returned the postcards said that they recycled. The second survey used a medium sample size and a random sampling method that specified going to schools in cities, small towns, farms, and so on. Thirty-seven percent of the schools said that they recycled.

The child being interviewed was shown descriptions of both surveys and was then asked to evaluate each survey. Sample interview questions addressing sampling method were: "How did each survey company decide what schools they were going to ask?"; "Can you see any advantages or disadvantages to either method?"; or "Do you think the schools who *are* recycling are more likely to be included in one survey or the other?" Sample interview questions addressing sample size were: "Is the number of

schools you ask important?"; "Did the survey companies ask enough schools?"; or "How do you know how many schools are enough schools?" Children were also asked how the surveys can have different conclusions and how they decided which (if any) one to believe.

Interview Procedure

Each child participated in 2 individual interviews that each lasted about 45 minutes. Interviews were semi-structured thereby allowing further investigation of individual issues that arose. The interviewer gave children positive feedback on all responses regardless of their accuracy. Children had ample time to complete the tasks. Interviewing took place during regular school hours in a small, quiet room near the classroom. All interviews were audio-taped and transcribed.

Study 1 Results and Discussion

Children's responses to a variety of interview questions in multiple scenarios in two contexts (school and out-of-school (recycling)) were analyzed to determine the major categories of children's thinking about sampling issues. Analyses identified patterns in the children's responses, and the following sections provide descriptions and supporting quotations for the major response categories for each of the study's questions: (a) How do children evaluate sampling methods, and (b) How do they draw conclusions from multiple surveys?[2]

Reliability. A sample of 69 quotations was selected for reliability coding. An additional coder sorted these quotations into the defined categories and intercoder agreement was 99%.

How do children evaluate different sampling methods?

Children were asked to evaluate a variety of sampling methods. Their responses suggested that sampling considerations are important to children. Furthermore, their evaluation rationales fell into four main categories that focused on: (1) potential for bias, (2) fairness, (3) practical issues, and (4) results. Children were not consistent in their use of these four types of rationales when evaluating sampling methods. Rather, all children used multiple types of evaluation rationales and sometimes children used more than one type of rationale when evaluating a specific sampling method in a single scenario.

1. Focus of Sampling Method Evaluations: Potential for Bias

Some children evaluated sampling methods by focusing on just what we want them to consider: the quality of the sample and the potential for bias. Some children accurately evaluated sampling methods by focusing on this potential for bias in the resulting sample. Other times, children seemed to focus on the potential for bias but inaccurately evaluated the sampling method. In other words, they made the wrong evaluation but for the right reason. The following sections describe how children used rationales based on potential for bias when evaluating all three types of sampling methods: (a) restricted, (b) self-selected, and (c) random sampling methods.

---

[2] Study 1 results do not provide frequency counts for the number of children using each response category. These counts would not have been meaningful given that the interview was semi-structured and not all children received the same opportunities to provide each response. However, it is important to note that in order to be considered a category, a certain type of response needed to be provided by more than one child. Study 2 results do provide frequency counts to explore the prevalence of each response category among a larger number of children.

Focus on potential for bias with restricted sampling. Children who focused on the potential for bias accurately evaluated restricted sampling methods negatively and discussed how restricted sampling methods are likely to produce samples that contain a preponderance of individuals with a particular opinion. In the following excerpt from the raffle scenario, the child is commenting on a restricted sampling method that only asked the friends of the person conducting the survey.

Lisa[3]: Some of these are not very good ideas.
Int: Ok, which ones?
Lisa: Like, well the first one where he asked 60 friends. Friends a lot of times are friends because they have the same opinions . . . So a lot of his friends are going to like one thing or the other. And so, it seems to me they mostly like getting raffle tickets [results were 75% would buy a raffle ticket]. . . It wasn't the best way. It could have been done better.

Focus on potential for bias with self-selected sampling. Some children focused on the potential for bias and *accurately* evaluated self-selected sampling methods negatively because individuals who chose to complete surveys were likely to have different opinions than those who did not choose to participate.

[Raffle scenario] Claire set up a booth outside the lunchroom and anyone who wanted to could stop by and fill out her survey. To advertise her survey she had a sign that said "WIN A SEGA". She stopped collecting surveys when she got 60 completed.

Carol: People that only really wanted to have that raffle would fill it out, and the people that didn't care would probably not.
Int: So is that good or bad?
Carol: So it's not good because people might just want a SEGA and say "oh, let's fill it out so that we can have a chance of winning."
Int: Ok. And so that wouldn't be a good way cause it might?
Carol: Only put people that want it.

Other children focused on the potential for bias but *inaccurately* evaluated self-selected sampling methods positively. They assumed that self-selected sampling methods would produce a good mixture of respondents because of the *absence* of specified sample restrictions, such as only selecting girls or one's friends.

[Raffle scenario] Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her.

Carol: I think Abby's was a good idea because . . . she got a variety of people.
Int: Is she sure she got a variety?
Carol: Well she probably got a variety because you weren't just giving them to a couple people and then giving it back to you. You were asking every kid, and whoever wanted to return it, they could.

These responses seemed to focus on the potential for bias yet missed the most probable source of bias with self-selected sampling methods: the idea that some types of individuals are more likely to participate than other types of individuals.

---

[3] All children's names are pseudonyms. In all excerpts, "Int" refers to the interviewer.

Children's lack of recognition of potential bias was particularly apparent in the recycling scenario, which included a self-selected sample of principals who returned postcards to indicate whether or not their school was recycling. For some children, this sampling method provided an obvious motive why principals may not want to participate in the survey:

> Karen: The majority of the schools didn't send back their postcards so it could mean they don't recycle and they're kind of embarrassed . . . so these people [ones who sent back postcards] might be the ones that were proud to show they recycled.

On the other hand, children who did not mention this potential bias were specifically asked "If you were a principal and your school did *not* recycle, do you think you'd send back your postcard?" Even when faced with this direct question, some children provided idealistic or altruistic responses rather than recognizing the motive for not participating.

> Amanda: I probably would, to be honest. And then I'd start a program to recycle.
>
> Melanie: Yeah . . . because they want information and if you just don't give it to them then they might spend their money the wrong way and so you might want to help them out a little.

This trust in the goodness of people's intentions[4] helped make self-selection bias difficult to detect for some children. They inaccurately evaluated self-selected sampling methods positively because, without any obvious restrictions, these methods should produce a mixture of respondents.

Focus on potential for bias with random sampling. Some children focused on the potential for bias and *accurately* evaluated random sampling methods positively because they asked a mixture of people. They reasoned that a mixture is important because different types of people have different opinions.

> Kyle wanted the same number of boys and girls and some kids from each grade. So he asked 5 boys and 5 girls from each grade to get his total of 60 kids.
>
> Lisa: That one [Kyle's method] looks pretty good cause that way he has a mixture of boys and girls and who are different ages.
> Int: Ok, and why is it important to have a mixture of boys and girls and different ages?
> Lisa: Well, because sometimes girls and boys can have different opinions on things and also one age might really like something, but an older age might think that was, you know, a terrible idea.

Furthermore, some children liked to specify the mixture and therefore preferred stratified random sampling (i.e., selecting 5 boys and 5 girls from each grade) to simple random sampling (i.e., picking 60 students' names out of a hat containing all 600 names) in the raffle scenario. Bill complained that with simple random sampling it was impossible to know what you were going to get: "could have a lot of variety and could have a little variety but it depends what names she picks."

When the mixture was not clearly specified, some children focused on the potential for bias but *inaccurately* evaluated the sampling method negatively. They did not like the uncertain or unknown

---

[4] In many surveys using self-selected sampling methods, individuals who do not participate are not purposely avoiding participation. Indifference is a major factor in low response rates for many surveys based on self-selection. However, these children were unable to detect the potential bias even when there was a strong motive for avoiding participation.

quality about who was going to get selected. Some children expressed their concern in more specific terms in that they felt more of a certain type of individual (with certain opinions) could be selected. Amanda suggested that with simple random sampling "you could get like all your friends, or all girls, or all boys, like all in first grade or something and everybody else has different opinions." These restrictions (e.g., all friends or all girls) were the basis of other (restricted) sampling methods and Amanda was not convinced that simple random sampling would protect against these biases. Children seemed to focus on the *possibility* of extreme outcomes without realizing that the *probability* of their occurrence was low.

Given their preference for stratification, did these children recognize that stratification variables should be related to the question being asked? Some children did demonstrate this link by distinguishing between good and bad stratification variables for a particular survey question. For example, in the raffle scenario, one child suggested that pet ownership is irrelevant but siblings (and their experiences with SEGA™) are not.

> Int: [asking about potential stratification variables] What about like whether they have a pet or whether they have an older brother or sister?
> Karen: I don't think the pet is a factor, but like with the older brother or sister they might have more experience with that [SEGA™] because their older brother or sister might like it. Like a first grader might not have one or something but their older brother or sister might and they might use that a lot.

In short, in some cases, children were able to justify their preference for stratified random sampling by identifying the potential links between the stratification variables and the individuals' responses to the survey question. In other cases, they were unable to articulate a reason for their preference. Rather, stratifying specified a mixture, and a good mixture was a worthy goal by itself, regardless of the relationship to the question being asked.

Summary of sampling method evaluations focused on bias. Some children made accurate evaluations by focusing on the potential for bias. With this focus, children were able to identify potential bias with restricted and self-selected sampling methods and to recognize a lack of potential bias with random sampling methods. With self-selected and random sampling methods, children sometimes focused on the potential for bias but made *inaccurate* evaluations. With self-selected methods, some children missed the most obvious source of bias (i.e., self-selection) and instead expected a mixed sample and consequently an unbiased sample. With random sampling methods, some children mistrusted the unknown nature of simple random sampling, rather than recognizing that there were no obvious sources of bias with this method.

2. Focus of Sampling Method Evaluations: Fairness

Rather than focusing on the potential for bias, some children evaluated sampling methods based on whether or not they were fair. These children were not thinking of fair in the probabilistic sense but rather in the affective sense of how the participants (or non-participants) felt about having the opportunity to participate in the survey. For example, when evaluating a sampling method of picking only one's friends, one child commented:

Andrea: That still wouldn't be fair. Because some people don't know him and maybe one of his friends knew this other person who was their friend and they would say, "Hey, but this person told me that you picked them and not me, how come?"

This perspective is based on the assumption that everyone wants to participate in a survey and everyone should have a chance to participate. Despite the apparent similarity between this goal and the idea of random sampling, there is an important difference. These children are not concerned that everyone has an *equal chance of being selected* so that the sample is not biased. Rather they are interested in how individuals *feel* when they are selected (or not selected) to participate.

Sometimes a fairness rationale supported accurate evaluations of restricted, self-selected, and random sampling methods. For example, Andrea (above) accurately concluded that a restricted sampling method was a bad idea but for the wrong reasons. She did not like the sampling method because the people who were left out would feel bad, not because the responses would be restricted and potentially biased by the people selected.

Other times, the fairness rationale led to inaccurate conclusions. This problem was particularly prevalent when children were evaluating self-selected sampling methods. With these methods, everyone initially has a chance to participate, and consequently, many children evaluated them as fair and good.

[Raffle scenario] Claire set up a booth outside the lunchroom and anyone who wanted to could stop by and fill out her survey. To advertise her survey she had a sign that said "WIN A SEGA". She stopped collecting surveys when she got 60 completed.

Int: Do you think that's [Claire's booth] a good idea?
Ed: Yeah because the people will choose if they want to. Or if they don't want to, they don't have to. Like if they wanted to do the survey they will, but if they would not want to, they don't have to -- so they're not pressuring anybody.

The fact that everyone had a chance to participate (i.e., their idea of fairness) was more important than the idea that some types of people (with particular opinions) were more likely to participate than others. This preference can be partially explained by the fact that some children did not recognize that certain types of people were more likely to participate than others.

A rationale focused on fairness is consistent with earlier findings by Jacobs and Lajoie (1994) and Schwartz and colleagues (1994). The salience of fairness is not particularly surprising given that children have numerous experiences with issues of fairness early in their lives (see, for example, Hunting (1991) for children's mathematical reasoning; Damon (1975) and Enright et al. (1984) for children's reasoning about justice issues in the social domain). Additionally, this over-reliance on fairness (at the expense of reasoning about inference) may not be restricted to children. Similar reasoning was found in a study of freshmen undergraduates enrolled in a course entitled "Reasoning in an Uncertain World" (H. P. Osana, personal communication, February 24, 1995). These findings underscore the utility of specifically exploring the fairness rationale and the importance of examining children's rationales in general when evaluating sampling methods -- they may make good decisions but for the *wrong* reasons.

3. Focus of Sampling Method Evaluations: Practical Issues

Some children evaluated sampling methods based on whether actually conducting the survey would be practical. For example, was the sampling method efficient, easy to do, confusing, or even possible?

> Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her.
>
> Bill: Well, she [Abby] asked every kid in the school and if she's on like 29, and then just a whole lot of people come in and give their surveys, she'll lose count. She might lose count and then start all over again! . . . So it's not very efficient.

Sometimes children focused exclusively on practical issues when evaluating a particular sampling method. Other times, they would mention practical concerns in conjunction with other explanations. Children used practical rationales when evaluating restricted, self-selected, and random sampling methods.

Despite this reliance on practical rationales, children were not always accurate in their evaluations of what sampling methods were, in fact, practical. For example, in the recycling scenario, Andrea suggested "I think I'd do it like the Question Center did [sending out postcards to all schools] but I'd say 'You have to send this back.'" Her solution to the potential bias from a self-selected sampling method was to make everyone respond without any consideration for the practical problems in doing so.

4. Focus of Sampling Method Evaluations: Results

Some children evaluated sampling methods based on the results of the survey. Children incorporated the results into their sampling method evaluations based on two criteria: (a) correspondence of the results with their expectations and (b) decisiveness of the results.

Correspondence of results with expectations. Some children evaluated the sampling method by whether the results agreed with what they expected the results to be. If the results corresponded with what was expected, then it was an appropriate sampling method because it got the "right results." Conversely, if the results did not confirm an expectation then it was an inappropriate sampling method. These expectations came from two sources.

First, some children evaluated the sampling method based on whether or not the results agreed with their *a priori* opinions on the topic. Consistent with the research on biased assimilation (Lord et al., 1979), some children evaluated sampling methods more critically when they disagreed with the results than when they agreed with them. The following excerpt illustrates this type of evaluation in the context of the recycling scenario. In this scenario, children were asked to compare Answers Inc.'s survey, in which they found 37% of the schools were recycling, with The Question Center's survey, in which they found 91% of the schools were recycling.

> Int: Do you believe one of the surveys more than the other or do you think both of them are good or both of them are bad?
> Ed: I think Answers Inco was one of the best because not many schools in Wisconsin are recycling.
> Int: How do you know that?

Ed: Because it just seems like it. Because not many schools, except one I really really know -- but that was like down in Baraboo or something and it recycles and it's got a bird planetarium and fish aquarium from all the money they saved up with their pop cans and everything. And then this school of course.
Int: Ok, because you just already know that most schools don't recycle, the bottom one found out what you thought it should find out?
Ed: Yeah.

This child's rationale shows little understanding of sampling issues. This rationale is in contrast to children who were able to separate their evaluations from their own opinions. In the following excerpt from the raffle scenario, a child is commenting on the stratified random sampling method, which found 30% of the students said they would not buy a raffle ticket.

Karen: I think the way they picked the same number of boys and girls in each grade was a good way . . . I don't know about those results though.
Int: Oh what do you think about those results?
Karen: Well they don't seem to really match what most of the kids I know would think cause I know a lot of kids like video games.

Second, some children evaluated sampling methods based on whether the results agreed with what they thought the survey was trying to prove. In this case, children did not see the purpose of surveys as a quest for accurate information but rather as an attempt to prove a particular outcome. For example, in the raffle scenario, these children thought the purpose of the survey was to prove that most children would buy a raffle ticket rather to find out how many children would buy a ticket[5].

Int: Does it make a difference how they decided who they were going to ask? Are some ways better than others? Are they all OK? Or are none of them OK?
Melanie: Well some of them are better than others because they got different percentages. Because Claire, she got who will buy raffle tickets -- 100% -- and who will not buy raffle tickets -- 0. So she might have gotten a way for the people to think that buying raffle tickets are better or something.
Int: Ok, so you think her way is a good way because she got 100%? [Melanie agrees]
Melanie: If that's what she's trying to do or something.

Melanie evaluated the sampling method positively because the results perfectly accomplished her perceived goal of the survey: showing that 100% of the students said they would buy raffle tickets.

Decisiveness of results. Some children evaluated sampling methods based on the decisiveness of the results. A decisive result such as 100% for one option and 0% for the other option was more useful than an indecisive result such as 50% and 50%. Therefore, the sampling method that produced the 100% result was better than the sampling method that produced the 50-50 split.

Because of the small number of items, it was impossible to systematize variations of decisive and indecisive results. Consequently, it was sometimes difficult to distinguish children's evaluations based on the decisiveness of the results from those corresponding with the children's expectations. For

[5] Sometimes the purpose of a survey *is* to prove or promote a particular outcome rather than collect accurate information. For example, companies that use surveys in their advertisements do not want accurate information unless it supports the product they are trying to promote. However, in these scenarios, the children were told explicitly that the surveys were trying to find accurate information.

15 BEST COPY AVAILABLE

example, in the raffle scenario, one survey found that 100% of the students would buy raffle tickets and 0% would not. In this case, the results were decisive *and* supported the purpose of the survey (i.e., to show that most students would buy raffle tickets).

However, responses to another survey suggest that the decisiveness of the results may be a consideration in its own right. In this raffle survey, 50% of the students said they would buy raffle tickets and 50% said they would not. Some children did not like this sampling method because of the difficulty of drawing a conclusion. As Lisa suggested "50-50's not going to decide for you." Consequently, sampling methods whose surveys produce decisive results may be evaluated more favorably than those whose surveys produce indecisive results.

Summary of How Children Evaluate Different Sampling Methods

Children's rationales when evaluating a variety of sampling methods focused on the potential for bias, fairness, practical issues or results. All children used more than one evaluation rationale during the interviews.

How do Children Draw Conclusions From Multiple Surveys?

The second question Study 1 investigated was whether children consider the quality of the surveys before drawing conclusions from multiple surveys. Children were asked to draw conclusions from multiple surveys with different results in a variety of scenarios. Children's responses fell into four categories: (1) considering survey quality, (2) aggregating all surveys regardless of their quality, (3) using their own opinions and ignoring all survey data, and (4) refusing to draw conclusions based on their own opinions or the survey data.

1. Drawing Conclusions from Multiple Surveys Based on Survey Quality

Some children first evaluated the quality of the surveys and then used the information from the surveys they thought were done well. The survey quality was judged on a variety of dimensions such as sample size and sampling method. For example, in the following excerpt from the raffle scenario, children evaluated whether the sampling methods were appropriate:

> Int: Do you think they would raise a lot of money with raffle tickets?
> Lisa: Well, I think they'd probably raise more doing something else because some of these [surveys] aren't done very well so you can't count on those. And the ones that are done well seem to be towards the end -- will not buy the raffle tickets.

In short, some children drew conclusions from one or more surveys but only after they had evaluated the quality of the surveys; they ignored information from surveys when they felt the surveys (their sampling method, sample size, etc.) were inappropriate.

In some cases, children were able to recognize that a survey might have both good and bad qualities as compared to another survey. The recycling scenario had children compare two surveys that juxtaposed a higher quality sampling method with a larger sample size. The first survey used a large sample size and a self-selected sampling method (sending out postcards), and the second survey used a medium sample size and a random sampling method that specified going to schools in cities, small towns, farms, etc. When asked to draw a conclusion about school recycling, some children focused on

sample size, some on sampling method, and some on both. In the following excerpt, Georgia recognizes the trade-offs between sampling method and sample size in this scenario.

> Georgia: They're both like the same because they [self-selected method] didn't make sure that they went to cities, small towns, and farms.
> Int: Ok, why is that important?
> Georgia: So they get a variety of things I guess . . . because a farm, maybe they wouldn't recycle as much.
> Int: Ok, so they might be different from what happened in the city? [Georgia agrees] So you like how they [random method] went to a lot of different places? [Georgia agrees]
> Georgia: And I like how they [self-selected method] asked a lot of schools.

2. Drawing Conclusions by Aggregating Multiple Surveys Regardless of Survey Quality

When presented with multiple surveys, some children aggregated all the available information regardless of the quality of the surveys. How they aggregated this information varied across scenarios. Sometimes children would add all the percentages for each response choice across surveys and then compare the totals. Other times, they computed an average for each response choice across surveys and then compare the averages.

In cases in which there were more than two surveys presented, children sometimes counted how many times response choice "A" won and compared that to how many times response choice "B" won. The following except illustrates this win-loss method with the raffle scenario:

> Int: If . . . you were the teacher of that class, and you had this information, do you think it would be a good idea at your school to sell raffle tickets -- do you think that would help you raise money to go to Great America?
> Randy: Hmm. No I don't think so.
> Int: No? Ok, how come?
> Randy: Because right here it says that more people don't want to buy a SEGA cause it's 10% want to and 90% don't want to and then the next is 30-70, so that's also a loss, and then 25-75, that's also a loss, then 100%-0% that's a big win, 50-50 is um OK, 90-10 is a win, 55-45 is OK and then 35-65 is a loss and 75-25 is a win so it's about even out . . . so they should try something else.

At first glance, it is reasonable to aggregate all information on a topic. However, children often aggregated information that they had previously identified as problematic. For example, Randy (from the most recent excerpt) negatively evaluated all of the restricted sampling methods, yet was willing to include their results in his overall conclusions.

3. Drawing Conclusions from Multiple Surveys Using Personal Opinions and Ignoring Survey Data

Some children did not evaluate the quality of the surveys nor aggregate the information without regard to quality. Rather they drew conclusions by ignoring the surveys and their data altogether. These children drew their conclusions based on other information, usually from personal experiences or opinions.

4. Refusing to Draw Conclusions from Multiple Surveys

Some children essentially refused to draw conclusions either from the data presented or their own opinions. They became indecisive when presented with conflicting results.

Int: Well what do you think of these surveys? . . .
Felisha: I don't know actually. I can't think of what I would do if there were two different answers.

Sometimes these children sought additional surveys to give more support to one conclusion or the other. Other times, they delegated the decision-making to other individuals such as authority figures.

Summary of Study 1

Study 1 was designed to identify response categories of children's conceptions about sampling and drawing conclusions from multiple surveys. Children were asked to evaluate a variety of sampling methods. Their responses suggested that sampling considerations are important, and their evaluation rationales fell into four main categories. First, there were some children who focused on the quality of the sample and the potential for bias. In most cases, this focus served children well, and they identified potential bias in restricted and self-selected methods. Furthermore, they recognized that there were no obvious sources of bias with random sampling methods. However, sometimes children focused on the potential for bias but inaccurately evaluated sampling methods because they missed the source of bias (as in self-selection) or mistrusted the unknown nature of samples (as in simple random sampling). Second, some children focused on issues of fairness to the participants and non-participants. Third, practical issues drove the evaluations of some children. Fourth, others relied on the decisiveness of results and/or the consistency of results with their expectations.

Children were also asked to draw conclusions when they were presented with multiple surveys. Their responses fell into four main categories. First, there were some children who evaluated the quality of the surveys before drawing a conclusion. Second, some children aggregated all the surveys regardless of their quality. Third, others ignored the survey data and relied on their own experiences or opinions to help them draw conclusions. Fourth, for some children, multiple surveys with multiple conclusions prevented them from even drawing a conclusion.

STUDY 2

Study 2 was designed (a) to confirm the response categories identified in Study 1 and (b) to determine the prevalence of the response categories in a larger sample of children. Rather than open-ended interviews, Study 2 used paper-and-pencil tasks to assess children's preferences for different responses that were *presented*. Analyses examined the range and frequencies of children's preferences by context and type of sampling method.

Selection of Participants

One-hundred-ten fifth graders from eight classes in three elementary schools in the same medium-sized city in Wisconsin completed two sets of paper-and-pencil tasks in November and December. Schools were chosen to be representative of this city, and these schools reflected the minority and income distributions of the city (31.6% minority and 29.2% free and reduced lunch). Teachers of these classes reported that they had not discussed sampling issues prior to the study.

Paper-and-Pencil Tasks

A series of paper-and-pencil tasks combined an open-ended and forced choice format. Children were presented with questions about sampling issues and were then asked to select responses from choices that represented the response categories identified in Study 1. All questions also included an open-ended response that allowed children to generate their own responses if they did not agree with the choices presented. Children were also not limited to a single response. They were told to circle any response that they agreed with and to put a star next to the response they agreed with most (see Table 1 for a sample question).

Table 1. Sample Question Format for Study 2 from the Raffle Scenario

Raffi asked 60 of his friends and he found that 75% said they would buy raffle tickets and 25% said they would not buy raffle tickets.

What do you think of Raffi's survey?
(Circle one) good bad I'm not sure

Here are some ideas that other kids had. Circle any of the ideas that you agree with. Put a star next to the idea you agree with most. If you don't agree with any of the ideas, circle the last choice and explain what you think.

a) I made my decision because it was easy to do. He just had to ask people he already knew.

b) I made my decision because his friends probably agree with him. So the survey doesn't tell you how the people who are not friends with Raffi think.

c) I made my decision because most of the kids said they would buy raffle tickets.

d) I made my decision because it's not nice to the people who are not his friends. They want to answer the survey too but they aren't allowed.

e) I made my decision because_____
_____

Scenarios. The two main scenarios from the Study 1 interviews (plus some additional scenarios) were used in the paper-and-pencil tasks. Additionally, the same three types of sampling methods (i.e., restricted, self-selected, and random) were included. Within each scenario, sample size was held constant and results for each survey were skewed to accentuate potential biases. Table 2 shows how the three types of sampling methods and their results were represented in Study 2.

Table 2. Study 2 Examples of Sampling Methods and Results

Type of Sampling Method	School Raffle Scenario (N=60)	Recycling Scenario (N=1500)
Restricted	Jake asked 60 kids at an after school meeting of the Games Club. (90% said they would buy tickets) Raffi asked 60 of his friends. (75% said they would buy tickets)	The second survey company asked the 1500 schools that attended the Wisconsin Earth Day Celebration. (90% said they recycled)
Self-Selected	Claire set up a booth outside the lunchroom and anyone who wanted to could stop by and fill out her survey. She stopped collecting surveys when she got 60 completed. (95% said they would buy tickets) Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her. (85% said they would buy tickets)	The first survey company sent postcards to every school in Wisconsin (3000 schools). (Of the 1500 schools that sent back their postcards, 91% said they recycled)
Random	Shannon got the names of all 600 kids in the school, put them in a hat, and pulled out 60 of them. (35% said they would buy tickets) Kyle put the names of all the first grade boys in one hat and the first grade girls in another hat. He pulled out 5 boys and 5 girls from each hat. He did the same thing for each grade until he had 5 boys and 5 girls from each grade. (30% said they would buy tickets)	The third survey company threw all the school names in a box and picked out the 1500 schools they were going to call. (28% said they recycled)

Questions for evaluating sampling methods. For each sampling method, children were asked to identify a rationale for their evaluation from a list of at least four potential evaluation rationales. These rationales corresponded with each of the four main categories of rationales that children used in their evaluations in Study 1: potential for bias, fairness, practical issues, and results. For example, response A in Table 1 illustrates an evaluation rationale focused on practical issues, B illustrates a focus on potential for bias, C illustrates a focus on results, and D illustrates a focus on fairness.

Self-selected and random sampling methods each had an additional category that corresponded to the rationales that focused on the potential for bias but made inaccurate evaluations because they either

missed the source of bias (as in self-selection) or mistrusted the unknown nature of samples (as in simple random sampling).

Order and length for evaluation rationales were varied across questions within each type of sampling method. Similarly, categories that could lead to positive and negative evaluations were varied. For example, a practical rationale could lead to a positive evaluation if the sampling method is easy to implement or to a negative evaluation if the method is difficult to implement.

Children in Study 2 were also asked how they would conduct a survey if they wanted to answer the questions the surveys were designed to answer. Children were asked to choose from the surveys presented or to provide a description of what they would do differently.

Questions for drawing conclusions from multiple surveys. Study 1 identified four means of drawing conclusions from multiple surveys. Children drew conclusions based on (1) the quality of the surveys, (2) aggregation of all the surveys regardless of survey quality, or (3) personal opinions. The fourth type of response occurred when children refused to draw *any* conclusions. In Study 2, for several scenarios, children were asked to draw a conclusion and identify how they drew that conclusion from a list of the four response categories of drawing conclusions (see Table 3). In some cases, there was more than one choice provided for a particular response category. The order of the response categories was varied across scenarios.

Table 3. Study 2 Example of Asking Children to Draw Conclusions from Multiple Surveys

Response Category	**What percent of kids in the whole school will buy a raffle ticket?** **Shannon pulled 60 names out of a hat** **35% will buy tickets** **65% will not buy tickets** **Claire set up a booth to collect 60 surveys** **95% will buy tickets** **5% will not buy tickets** **Jake asked 60 kids in the Games Club** **90% will buy tickets;** **10% will not buy tickets**
Consider survey quality	I thought Shannon's survey was the only one that was done well so I ignored the other 2 surveys and used Shannon's results. She found that 35% said they would buy a raffle ticket. I thought Claire's survey was the only one that was done well so I ignored the other 2 surveys and used Claire's results. She found that 95% said they would buy a raffle ticket. I thought Jake's survey was the only one that was done well so I ignored the other 2 surveys and used Jake's results. He found that 90% said they would buy a raffle ticket.
Aggregate all surveys	I took the average of the 3 surveys. The average of the kids who said they would buy a raffle ticket is 73%. I looked at the surveys and saw that 2 surveys (Claire's and Jake's) said that most kids would buy a raffle ticket and only 1 (Shannon's) said that most kids would not buy a raffle ticket. I used the information from the 2 surveys that agreed so 90%-95% said they would buy a raffle ticket.
Use personal opinion and ignore all survey data	I just knew that most kids like SEGA and would buy a raffle ticket so I picked a high percent. ____ percent of kids will buy a raffle ticket. I just knew that most kids would not buy a raffle ticket so I picked a low percent. ____ percent of kids will buy a raffle ticket.
Refuse to draw conclusions	I don't know because they got different results.

Reliability

The same person who performed reliability coding for Study 1 coded the choices for the Study 2 paper-and-pencil tasks. She identified each choice as one of the response categories identified in Study 1. Intercoder agreement was 100%.

Procedure

Children individually completed the paper-and-pencil tasks. The same researcher administered the tasks in all eight classes during regular class time. The researcher read the tasks aloud to the whole class, and the children had ample time to complete the tasks. The tasks were administered in 2 parts, with each part requiring approximately 45 minutes to complete.

Study 2 Results and Discussion

Children's responses to the paper-and-pencil tasks were analyzed to confirm and determine the prevalence of the response categories identified in Study 1. Study 2 results provide additional evidence for the Study 1 identification of children's conceptions and the following sections indicate the frequencies of the response categories for how children evaluate sampling methods and draw conclusions from multiple surveys.

It is important to note that because of the format of the paper-and-pencil tasks, Study 2 assessed children's thinking by the responses they *recognized* rather than *recalled*. However, children used the entire range of Study 1 responses as each rationale was selected at least once for each question.[6] Furthermore, children used the open-ended responses sparingly, but 76.4% of them did use open-ended responses at least once. (Open-ended responses were provided as an option for each question in case children did not like any of the choices presented.) The percentage of children providing an open-ended response ranged from 2.7% to 30.0% across questions (M=10.5% of the children per question)[7]. Some of the open-ended responses were recoded into existing categories if the open-ended response was essentially a restatement of one of the listed categories and/or a slight variation on the same theme (M=2 recoded responses per question). This range suggests that children were taking the tasks seriously and were able to distinguish among the responses presented.

How Do Children Evaluate Sampling Methods?

The majority of the paper-and-pencil tasks addressed children's thinking when evaluating different sampling methods. Children were asked to provide an initial good/bad evaluation of nine different surveys and then identify a rationale for their thinking (see Table 1). After evaluating all the surveys related to a particular scenario, children were then asked to select their favorite sampling method or to describe the method they would use if they were conducting the survey themselves. The following

---

[6] Although children were asked to circle any response they agreed with and to put a star next to the idea they agreed with most, they most often agreed with a single response (M=70.9% of the children per question). Sometimes they chose two responses (M=21.8% of the children per question) and less frequently, they chose more than two responses (M=6.4% of the children per question). Given the predominance of single responses, it is not surprising that the patterns for the starred (*) responses were similar to the patterns when all circled responses were considered. Consequently, the following sections describe the results only in terms of the children's favorite, starred responses.

[7] The open-ended responses are not discussed further in this paper given that the agreement among students was not frequent enough to warrant the formation of new categories.

BEST COPY AVAILABLE

sections describe the results of children's initial good/bad evaluations and their rationales. Responses are examined overall, by context (school or recycling), and by type of sampling method (restricted, self-selected, or random). Finally, children's favorite sampling methods are explored.

Initial Evaluations of Sampling Methods

Children were asked to evaluate each survey as "good," "bad," or "I'm not sure" before providing an evaluation rationale. This question was designed to assess children's initial reactions to the survey before providing them with a list of evaluation rationales or ways of thinking about the survey. Overall, children's responses approximately fell into thirds. Thirty-nine percent of the responses were "good," 33.9% were "bad," and 25.7% were "I'm not sure."

Children also had different preferences for different types of sampling methods. Each sampling method was used three times (twice in the school raffle scenario and once in the recycling scenario). In the school raffle scenario, children negatively evaluated restricted sampling methods (M=56.8%) while in the recycling scenario, 58.2% positively evaluated the restricted sampling method. Perhaps this differentiation by context was an indication that bias was more immediately recognizable in the school context. Conversely, children consistently evaluated self-selected sampling methods positively (M=53.0%). This pattern held for all three surveys (and both contexts) involving self-selected sampling methods.

For the three surveys based on random sampling methods, each was evaluated somewhat differently. For the simple random sampling method in the recycling context, 13.6% of the children positively evaluated the survey. For the simple random sampling method in the school context, the children were spread evenly across the three choices of "good," "bad," and "I'm not sure." Finally, for the stratified random sampling method in the school context, 54.5% of the children positively evaluated the survey. These results support the Study 1 finding that children preferred stratified random sampling (i.e., specifying the mixture) to simple random sampling. In short, for random sampling methods, both context and type of random sampling method seemed to influence initial evaluations. However, given the small number of surveys evaluated, it is impossible to know if some of these variations are due only to specific components of the individual survey descriptions.

Rationales for Evaluations of Sampling Methods

Children were asked to select a rationale supporting their initial evaluations. The results confirmed that the evaluation rationale categories identified in Study 1 were all viable rationales as children used the full range of evaluation rationales for every question. Children most frequently selected evaluation rationales focused on the potential for bias and fairness (see Table 4). When considering children's evaluation rationales for all nine scenarios, only 6% of the children never focused on the potential for bias and 13.6% never focused on fairness.

Table 4. Percentage of Children Selecting Each Evaluation Rationale Category Across Nine Surveys

Rationale Category	Mean % (*SD*)	Min %	Max %
(Accurate[a]) Potential for Bias	34.0 (8.6)	18.2	45.5
(Inaccurate) Potential for Bias	15.6 (7.4)	5.5	28.2
Fairness	23.0 (9.6)	10.9	37.3
Practical Issues	12.1 (4.0)	3.6	17.3
Results	12.2 (7.3)	1.8	24.5

[a] The results make a distinction between accurate or inaccurate evaluations based on the potential for bias. The (accurate) potential for bias refers to the evaluation rationales that identified potential sources of bias (as in restricted and self-selected sampling) or recognized that they were missing (as in random sampling). The (inaccurate) potential for bias refers to the evaluation rationales that focused on the potential for bias but either missed the source of bias (as in self-selection) or mistrusted the unknown nature of samples (as in simple random sampling). If not specified, the potential for bias refers to the (accurate) potential for bias -- the types of criteria that a sophisticated evaluation of sampling method might use.

Evaluation rationale frequencies by context. The mean percentage of children selecting evaluation rationale categories varied somewhat by context. The fairness rationale was approximately twice as prevalent in the school context (*M*=27.7%, *SD*=8.4%) as in the recycling context (*M*=13.6%, *SD*=2.0%). Perhaps the idea of being selected or not selected to participate in a *school* survey was more personal and relevant to the children. Consequently, they would be more likely to have an affective response to the sampling method.

Evaluation rationales focused on results were almost three times as prevalent in the recycling context (*M*=21.2%, *SD*=3.0%) as in the school context (*M*=7.7%, *SD*=3.9%). Perhaps children focused more on results in the recycling scenario because it was more evident what those results should be. The recycling scenario described surveys to determine if schools were recycling. Even though the purpose of the surveys was to acquire accurate information, the desirable ultimate outcome was clear: to have most schools recycle. On the other hand, the school context described surveys to determine whether to raffle a SEGA™. It was more debatable whether raffling a SEGA™ was a good idea or not.

Evaluation rationale frequencies by type of sampling method. The mean percentage of children selecting evaluation rationale categories also varied by type of sampling method. With restricted and random sampling methods, children focused most often on the (accurate) potential for bias. For restricted sampling methods, the mean of children selecting rationales focused on potential for bias was 42.7% (*SD*=3.9%) while the next closest was the fairness rationale category with a mean of 26.7%

(*SD*=7.3%). For random sampling methods, the mean of children selecting rationales focused on potential for bias was 32.4% (*SD*=6.3%) while the next closest was the fairness rationale category with a mean of 16.1% (*SD*=6.0%).

With self-selected sampling methods, children's evaluation rationales were more evenly split between focusing on the (accurate) potential for bias and fairness (*M*=27.0%, *SD*=6.2% for potential for bias; *M*=26.4%, *SD*=10.7% for fairness). This result supports the Study 1 finding that children had difficulty identifying the potential bias with self-selected sampling methods. Instead, self-selected sampling methods were affectively appealing because everyone initially had an opportunity to participate; no one was excluded due to stated restrictions.

Favorite Sampling Method

For both the recycling and raffle scenarios, children were asked to indicate their preferred sampling method if they were conducting the survey themselves. In the raffle scenario, on average, 3.6% preferred the restricted sampling method, 39.1% the self-selected sampling method, and 42.8% the random sampling method. In the recycling scenario, 12.0% preferred the restricted sampling method, 40.0% the self-selected sampling method, and 16.4% the random sampling method.

In both contexts, children liked the self-selected sampling methods. This result is consistent with the Study 1 findings and the Study 2 results of children's initial evaluations. This partiality toward self-selected sampling sometimes existed even when children were able to recognize the potential bias in self-selection. For the self-selected sampling method in the recycling context, 36.4% of the children who recognized the bias when evaluating the survey still picked this method as their favorite method. In the school context, this mean percentage was 9.3%.

In addition to a preference for self-selected methods, children also liked the random sampling methods in the school context. However, this preference was due almost exclusively to their fondness for stratified random sampling methods. In the school context, 5.5% of the children selected the simple random sampling method as their favorite method while 37.3% selected the stratified random sampling method. This distinction again supports the Study 1 finding that children liked specifying the stratifications to gain some control over the sample composition.

How Do Children Draw Conclusions from Multiple Surveys?

There were two types of scenarios in which children were asked to draw conclusions from multiple surveys: two scenarios in which the surveys varied on the sampling methods used, and two scenarios in which the surveys varied on both the sampling methods and sample sizes used. Children drew conclusions somewhat differently in each type of scenario (see Table 5).

Table 5. Percentage of Children in Each Response Category When Asked to Draw Conclusions from Multiple Surveys

	Scenarios Varying Sampling Method		Scenarios Varying Sampling Method and Sample Size[a]	
Response Category	School Context	Recycling Context	School Context	Recycling Context
Consider survey quality	26.4	12.7	49.2	62.3
Aggregate all surveys	41.8	46.3	15.4	1.6
Use personal opinions	8.2	20.9	18.5	21.3
Refuse to draw conclusions	15.5	15.5	13.8	13.1

[a] Due to time constraints, not all 110 children responded to this type of scenario. Sixty-one children in the school context and 65 children in the recycling context responded to this type of scenario.

When drawing conclusions from multiple surveys in most scenarios, the majority of children considered survey quality or aggregated all surveys regardless of their quality. When both sampling method and sample size varied, a larger percentage of children considered survey quality. In these scenarios, a survey with a random sampling method and moderate sample size was juxtaposed with a restricted sampling method and a large sample size. Perhaps this juxtaposition highlighted issues of quality. Furthermore, children could focus on survey quality by preferring information from the survey with the random sampling method or the survey with the larger sample size. This preference varied by context. In the school context, 32.3% of the children focused on sampling method while 12.3% focused on sample size. In the recycling context, 27.9% focused on sampling method while 34.4% focused on sample size. Perhaps the larger numbers (in an absolute sense) in the recycling context drew more attention to sample size.

Relationship Between Drawing Conclusions from Multiple Surveys and Evaluating a Single Sampling Method

For many children, the task of drawing conclusions from multiple surveys seemed to be somewhat unrelated to the task of evaluating a particular sampling method. Specifically, children who were capable of identifying potential bias with a particular sampling method in a single survey did not always use this information when drawing conclusions from multiple surveys. For example, to draw a conclusion in the recycling context, 72 children aggregated the three surveys, each based on different sampling methods (i.e., restricted, self-selected, and random sampling methods). However, when evaluating each survey individually, 36.1% of these children had identified potential bias with the restricted sampling method and 36.1% of them had identified potential bias with the self-selected

sampling method. Similarly, when drawing a conclusion in the school context, 46 children aggregated three surveys even though 47.8% of them had previously identified potential bias with the restricted sampling method and 26.1% had identified potential bias with the self-selected sampling method. In short, children who were capable of identifying potential bias with a particular sampling method did not always use this information when drawing conclusions from multiple surveys. This result is consistent with the Study 1 findings.

Summary of Study 2

Study 2 confirmed the response categories for evaluating sampling methods identified in Study 1. Children evaluated sampling methods by focusing on potential for bias, fairness, practical issues, or results. Most children focused on the potential for bias or fairness. Furthermore, the ability to evaluate sampling methods was not restricted to a small group of "smart" children. Ninety-four percent of the children correctly used a sophisticated evaluation rationale (a focus on the potential for bias) at least once during the paper-and-pencil tasks. On average, 34% of the children accurately focused on the potential for bias per question. Consequently, the topic of evaluating sampling methods appears appropriate for children at this age. Some evaluation rationales varied depending on the context and type of sampling method. Children used fairness rationales more often in school contexts and rationales focused on results more often in recycling contexts.

Restricted sampling methods proved to be the easiest sampling methods to evaluate accurately. In general, children initially evaluated these sampling methods negatively and were more likely to focus on the potential for bias in their evaluation rationales as compared to their evaluation rationales for other sampling methods. Furthermore, they were unlikely to pick restricted sampling methods as their favorite sampling method. It is interesting to note that although 58.2% of the children initially positively evaluated the restricted sampling method in the recycling scenario, their evaluation became more negative once the rationales were presented. For this survey, 37.3% recognized the bias [the most frequently chosen rationale for this survey evaluation] and only 3.6% of the children selected this method as their favorite sampling method. Perhaps children's ability to successfully evaluate restricted sampling methods was due to the fact that these methods actually stated their restrictions, thereby highlighting the potential for bias.

Self-selected sampling methods proved more difficult to evaluate successfully. In general, children initially evaluated these sampling methods positively and about 40% of the children identified self-selected sampling methods as their favorite sampling method in both contexts. This fondness for self-selection may be attributed to children's focus on the perceived fairness inherent in self-selection. Given that everyone initially has a chance to participate, self-selection satisfied children's sense of equity but often masked the potential bias due to individuals selecting whether or not to participate. On average, 26.4% of the children selected fairness rationales when evaluating self-selected sampling methods. However, despite the distraction of fairness, 27.0% of the children on average were able to successfully identify the potential bias with self-selected sampling. Consequently, discussion of the potential bias inherent in self-selection is not beyond the abilities of children at this age.

Random sampling methods elicited more mixed responses than did other types of sampling methods. In general, children liked stratified random sampling, as evidenced by their positive initial evaluations and selection of evaluation rationales indicating lack of obvious bias. Furthermore, when asked to select their favorite sampling method (out of six presented methods and the option of specifying their own method), 37.3% of the children selected the stratified random sampling method. In contrast, children did not like simple random sampling, as evidenced by their negative initial evaluations, their unwillingness to select simple random sampling methods as their favorite method, and their selection of evaluation rationales that indicated their mistrust of the unknown nature of random samples.

Study 2 also confirmed the response categories for drawing conclusions from multiple surveys identified in Study 1. Children drew conclusions by considering survey quality, aggregating all surveys regardless of their quality, relying on their personal opinions, or refusing to draw a conclusion at all. In both contexts, children most often considered survey quality or aggregated all surveys regardless of their quality. There was a variation by context in terms of whether children who considered survey quality preferred a better sampling method or a larger sample size. Almost three times as many children focused on the larger sample size in the recycling scenario than in the school scenario. Given the small number of tasks, it was difficult to assess children's consistency in drawing conclusions across scenarios. Perhaps most disconcerting was that children often ignored sampling quality when drawing conclusions from multiple surveys even when they could identify potential bias with individual sampling methods.

INSTRUCTIONAL IMPLICATIONS

The NCTM *Standards* recommend exploration of statistical sampling issues in grades 5-8. However, in the past, sampling with surveys has been an atypical topic for the elementary school curriculum. This project showed that not only have upper elementary children had experiences with surveys outside of school, but they also have developed understandings about some of the key issues related to interpreting and evaluating surveys. Therefore, discussion of these topics is not beyond the abilities of children at this age and instruction should capitalize on their informal knowledge.

The findings from this project should provide teachers with information about children's informal conceptions and misconceptions of sampling issues, thereby giving them a more solid starting point for instruction. This project categorized children's thinking when evaluating sampling methods and drawing conclusions from multiple surveys. Children's thinking varied somewhat by context and type of sampling method. Unfortunately, even when children were able to effectively evaluate sampling methods, they did not always consider survey quality when drawing conclusions from multiple surveys.

Furthermore, this project should alert teachers to the idea that interpreting and evaluating statistics (in addition to collecting, organizing, and describing data) is a potential context for examining statistical issues. Finally, this project's findings should encourage teachers to probe for children's reasoning behind their evaluations because they may make correct decisions for the wrong reasons. The following sections highlight specific instructional implications suggested by this project: (a) increasing children's

exposure to sampling, (b) selecting effective surveys to include in instruction, and (c) increasing children's consideration of survey quality when drawing conclusions from multiple surveys.

Increasing Children's Exposure to Sampling

Children need experiences discussing issues of sampling. While elementary school children often conduct surveys, these surveys rarely involve sampling, or, if they do, issues of inference are often ignored. For example, the Study 1 children reported that when they conducted surveys, they asked everyone in their class. Similarly, the Study 1 school used a school-wide survey to measure student attitudes toward school policies and personnel, but even this survey was really a census of the entire school. These situations are not unusual given that the layperson's use of the term "survey" often means getting a count of people's opinions about some topic regardless of whether one is asking a sample or the population. For example, one could survey a class to see how many children want to go on a field trip to the zoo; everyone would be asked whether or not they want to go to the zoo. Consequently, children need experiences discussing sampling issues, such as who and how many to sample. They also need to explore how these decisions may affect the quality of the results. Discussions could focus on surveys that the children conducted and/or surveys that others produced.

Perhaps instruction on statistical sampling could build on children's informal conceptions of nonstatistical sampling. For example, when asked to define a sample in the Study 1 interviews, children did not discuss statistical samples. Rather they mentioned food samples at the grocery store, product samples that arrived in the mail, or examples of things such as student writing. Nonetheless, some of the students' definitions included some of the elements of statistical samples: (a) the sample is part of the whole and (b) the smaller part gives you an idea of the whole. Specifically, Melanie suggested that a sample is "a piece of something whole -- it's like a peek." Georgia identified a sample as "a piece of food or carpet that gives you an idea of what the real thing is."

At the end of the interviews, after discussing many surveys, the interviewer defined a "sample" in surveys as the group of people who answers the questions when it would take too long to ask everyone. Children were then asked how that definition was similar to their definition of sample. Children were insightful about the similarities. For example, Melanie realized that the sample is part of the whole and provides some information, but she missed the idea that the sample should represent the whole. She stated that "it's like a part of something, like a part of a lot of people. It's like a peek of what the people think -- not everybody though." Georgia added the idea of representativeness to her explanation by stating that it's "just the same thing -- you just take like a group of people that represents the whole thing that you're asking." Consequently, perhaps instruction should explore children's understanding of nonstatistical sampling in order to build connections to statistical sampling in the context of surveys.

Selecting Effective Surveys for Instruction

The selection of surveys for instruction is non-trivial. There is a trade-off between discussing many surveys and discussing a smaller number of surveys in depth. Furthermore, the results of these two studies suggest that children may respond differently to certain contexts and types of sampling methods. Consequently, teachers must make careful instructional decisions about what surveys to use to foster

children's understanding. The following sections describe what the results of this project suggest when selecting (a) the context and (b) the sampling methods of surveys.

Selecting Survey Contexts

An elementary school curriculum is perhaps most likely to include surveys in school contexts, since these contexts are familiar to children. However, children responded somewhat differently to school and out-of-school contexts. Children more often used the fairness rationale when evaluating sampling methods in the school context than in the recycling context. Perhaps the familiarity of the context encouraged an affective reaction and the use of the fairness rationale (to the exclusion of the desired focus on potential for bias). Additionally, although not examined in this project, surveys that sample *within* a child's specific school may arouse other affective responses. For example, the child may conclude that a survey is inappropriate if the sample did not include him or her. Consequently, it may also be important for instruction to include out-of-school contexts in order to evoke alternatives to affective responses to surveys.

Selecting Sampling Methods

It is important that children learn how to focus on the potential for bias when interpreting and evaluating surveys based on all three types of sampling methods: restricted, self-selected, and random. However, children responded somewhat differently to each type of sampling method. The following sections provide suggestions for (a) increasing children's recognition of potential bias, (b) reducing their mistrust of simple random sampling, and (c) extending their acceptance of stratified random sampling. The final section addresses the teachers' decisions regarding the influence of results on children's evaluations.

Increasing children's recognition of potential bias. Restricted sampling methods (especially in school contexts) seemed to be the methods with which children were most easily able to recognize potential bias. Perhaps instruction could start with restricted sampling methods and gradually introduce more subtle sources of potential bias, such as self-selection. In the school context, recognizing the potential bias in self-selection proved more difficult because of the children's reliance on the fairness rationale. Furthermore, even when children did recognize the potential bias, they often continued to evaluate self-selected sampling methods positively. It seemed that the initial equal opportunity for participation (i.e., fairness) was more important than the potential for bias in the resulting sample. Rather than totally dismissing the fairness rationale, perhaps instruction could start with its basic premise: the idea that everyone should have an equal opportunity to participate. Instruction would then need to help children extend their understanding of this idea toward a realization that the importance of equal opportunity is to minimize the potential for bias in the resulting sample (rather than to minimize the negative feelings of the non-participants).

When children demonstrated non-normative reasoning, it may have been due to the mismatch between their interpretation of the tasks and the researcher's intended tasks. Newman, Griffin, & Cole (1989) caution that the researcher's tasks are not always appropriated by the child. For example, children's use of the fairness rationale raises the question of whether children interpreted the task of

evaluating sampling methods differently than was intended by the researcher. In the Study 1 interviews, children were asked to evaluate sampling methods based on which methods would produce the most accurate information (presumably from the most representative sample). In fact, before evaluating any sampling method, the interviewer and child discussed the purpose of the survey and what consequences might occur if the survey provided inaccurate information. However, children using the fairness rationale did not seem to have the goal of identifying sampling methods that would yield accurate information. Rather, their task seemed to be making everyone feel included.

The following excerpt from the raffle scenario illustrates how the researcher's task can be unrelated to the child's task.

> Int: What about Raffi asking his 60 friends, is that a good way to do it?
> Felisha: Well, I'm not sure because then everyone wants to be his friend because they want to have a chance in the survey. So I think that's OK but I would say that most people, when he does his survey, won't get a chance because of Raffi's friends. Because some people wanted to be in the survey and weren't Raffi's friends -- well, they won't have a chance.
> Int: Alright, so do you think it would affect -- if he just asked his friends -- do you think it would affect the information he gets?
> Felisha: No.

After the child uses a fairness rationale, the interviewer restated the intended task of acquiring accurate information. However, this goal clarification did not cause the child to rethink her evaluation.

Similar to the fairness rationale, children who focused on practical issues or results may have been interpreting the task of evaluating sampling methods differently than was intended by the researcher. Consequently, in order to increase children's understanding of sampling issues, teachers may need to negotiate an understanding of the task with children. This negotiation would provide a means of helping children appropriate the *intended* task.

Reducing children's mistrust of simple random sampling. Children consistently distrusted simple random sampling. Interestingly, many elementary classrooms contain multiple applications of simple random sampling. For example, clips or popsicle sticks with children's names are often randomly selected to determine seating arrangements, assign jobs, and so on. In the Study 1 interviews, when these examples were mentioned as types of random sampling, children usually evaluated these procedures positively (as opposed to their negative evaluations of simple random sampling in the interview tasks). Perhaps children need even more experiences with simple random sampling in real contexts, or perhaps the experiences that they have already had need to be linked *explicitly* to the concept of simple random sampling.

Extending children's acceptance of stratified random sampling. Children consistently liked stratified random sampling, although it was not always clear that they liked stratification as a means of increasing the representativeness of the sample. Consistent with Schwartz et al.'s (1994) findings, some children seemed more concerned that individuals of each type were included. This alternative focus was particularly apparent in scenarios in which the categories of the stratification variables existed in unequal proportions in the population. In the following excerpt, Lisa does not want to use a sample that

reflects the population composition in terms of gender. Rather she is more concerned that boys and girls are equally represented.

> [after Lisa had suggested a differential preference for toys by gender]
> Int: What if you were having a party and . . . you were going to have 200 kids that were coming and 150 of them were going to be girls and 50 were going to be boys. . . . And you wanted to know what kind of toys to bring for people to play with. So you were going to pick out some kids to question, say 20 kids. What would be the best way to do that -- to get the best estimate of what kind of toys you should bring?
> Lisa: Don't do it randomly.
> Int: Don't do it randomly? Why not?
> Lisa: Cause you're gonna get more girls if there are three times as many girls than there are boys and you're gonna get more girls. And they'll probably want one certain thing and the boys might want a different thing.

Children's difficulties with unequal proportions were also seen in response to the election survey scenario in which some children suggested surveying equal proportions of Republicans, Democrats, and Independents despite their unequal proportions in the population. In addition, the results to a Study 2 question about children's conceptions of fairness in a non-survey context suggest that children's preference for stratified random sampling (over simple random sampling) may be linked to their statistically non-normative conceptions of fairness. When asked how to select 6 children *fairly* for a field trip (out of a class of 20 girls and 10 boys), 83.0%[8] felt it would be more fair to pick 3 girls' and 3 boys' names from separate hats rather than to put all the names in a single hat and pull out 6 names. Furthermore, 48.7% of the children who wanted separate hats recognized that separate hats would mean boys had a better chance of going on the field trip. Children's conceptions of fairness seemed to be linked more strongly to getting every group equally represented rather than giving everyone an equal chance probabilistically or reflecting the population composition. To extend children's understanding of stratified random sampling, instruction needs to include both examples of stratifications based on unequal proportions in the population and tasks that require a reflection of the population composition. Discussions need to probe decisions about what stratification variables to use and the number of individuals to sample in each category of stratification.

Understanding the influence of results. When evaluating sampling methods, some children focused on the results. Specifically, children's evaluations were influenced by the decisiveness of the results and the correspondence of the results with their expectations. Consequently, teachers need (a) to decide whether to include results in the instructional tasks, and (b) if results are included, decide what specific results to use.

First, teachers need to decide whether to include results in the instructional task. Different issues may arise when children are interpreting and evaluating others' surveys rather than conducting a survey themselves. While making sense of results is eventually a component of both tasks, the timeline for decision-making is different. When conducting a survey, children must make decisions about sampling

---

[8] This percentage is out of 95 children as one class was excluded. This class had discussed the questions prior to recording their responses.

method and sample size before making sense of the results. When interpreting and evaluating surveys that others have created, the sampling method, sample size, and results must be considered simultaneously. Furthermore, this project did find that children sometimes focused on the results (instead of the desired focus on potential for bias) when evaluating sampling methods in surveys that others produced. Consequently, instruction needs to recognize that while the two tasks of evaluating and conducting surveys overlap considerably, they also address some different issues.

Second, if teachers decide to include results in the instructional task, they must decide what specific results they want to use. Further studies need to explore these issues more systematically, but this project suggests that some children respond differently to decisive and indecisive results. Similarly, some children respond differently to results that agree with their *a priori* opinions as opposed to those that disagree. Consequently, teachers need to consider carefully the actual results included in an instructional task.

Increasing Consideration of Survey Quality When Drawing Conclusions from Multiple Surveys

All of the previous instructional implications addressed children's ability to evaluate survey quality. However, this project found that even when children were able to identify potential bias for individual surveys, they often ignored survey quality when drawing conclusions from multiple surveys. Instruction needs to address this oversight. Perhaps this oversight is related to Schwartz et al.'s (in press) and Jacobs' (1993) findings that children do not always understand the importance of representative sampling. Consequently, in order to increase the consideration of survey quality when drawing conclusions from multiple surveys, instruction would need to address not only how to evaluate surveys effectively but also *why* it is important to do so.

Final Thoughts

The findings from this project provide some suggestions for issues that instruction should address. However, perhaps an even larger question is how to begin to incorporate the topics of sampling issues into the curriculum. Perhaps sampling issues could be addressed as part of instruction on surveys. This instruction would address both sampling issues and other issues affecting survey quality (e.g., question wording, question format, and interviewer bias). Or perhaps sampling issues could be addressed through instruction in both survey and non-survey contexts (e.g., product testing or counting unseen populations such as whales). Or perhaps sampling issues could be linked to instruction on mathematical concepts such as percentages or proportional reasoning.

Regardless of how sampling issues are incorporated into the curriculum, many teachers will also need to struggle with their own understanding of these issues. Given that the average adult is generally a poor statistician and that statistics has been an atypical topic for pre-college level education, a staff development component will be required for the implementation of any of the NCTM *Standards* in statistics. Perhaps the findings from this project will also be useful in helping adults develop a better understanding of sampling issues, as the categories of responses identified in this project may not be limited to children.

REFERENCES

Asher, H. (1988). *Polling and the public: What every citizen should know.* Washington, D.C.: Congressional Quarterly.

Carpenter, T. P. & Fennema, E. (1992). Cognitively Guided Instruction: Building on the knowledge of students and teachers. In W. Secada (Ed.), *Curriculum reform: The case of mathematics in the United States.* Special Issue of *International Journal of Educational Research* (pp. 457-470). Elmsford, NY: Pergamon Press.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22* (1), 3-29.

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. *Review of Research in Education, 16,* 3-56.

Crossen, C. (1994). *Tainted truth: The manipulation of fact in America.* New York: Simon & Schuster.

Day, R. P., Webb, N. L., Nabate, M., & Romberg, T. A. (1987). *Evaluation of quantitative literacy series: Exploring data and exploring probability* (Program Report 87-5). University of Wisconsin-Madison, Wisconsin Center for Education Research.

Damon, W. (1975). Early conceptions of positive justice as related to the development of logical operations. *Child Development, 46,* 301-312.

de Lange, J., Burrill, G., Romberg, T., & van Reeuwijk, M. (1993). *Learning and testing mathematics in context: The case: Data visualization.* Scotts Valley, CA: Wings for Learning.

de Lange, J., & Verhage, H. (1992). *Data Visualization* (2nd version). Scotts Valley, CA: WINGS for Learning.

Dewdney, A. K. (1993). *200% of nothing: An eye-opening tour through the twists and turns of math abuse and innumeracy.* New York: John Wiley & Sons.

Empson, S. (1995, April). *Equal sharing and shared meaning: The development of fraction concepts in a first-grade classroom.* Paper to be presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Enright, R. D., Bjerstedt, A., Enright, W. F., Levy, V. M. Jr., Lapsley, D. K., Buss, R. R., Harwell, M., & Zindler, M. (1984). Distributive justice development: Cross-cultural, contextual, and longitudinal evaluations. *Child Development, 55,* 1737-1751.

Gal, I., Rothschild, K., & Wagner, D. A. (1989, April). *Which group is better? The development of statistical reasoning in elementary school children.* Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63* (1), 25-34.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27* (3), 337-364.

Hiebert, J. & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York: Macmillan.

Hunt, E. & Minstrell, J. (1994). A Cognitive Approach to the Teaching of Physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 51-74). Cambridge, MA: MIT Press.

Hunting, R. P. (1991). The social origins of pre-fraction knowledge in three year olds. In R. P. Hunting & G. Davis (Eds.), *Early fraction learning* (pp. 55-72). New York: Springer-Verlag.

Jacobs, V. (1993). *Stochastics in middle school: An exploration of students' informal knowledge*. Unpublished master's thesis. University of Wisconsin-Madison.

Jacobs, V. & Lajoie, S. (1994, April). *Statistics in middle school: An exploration of students' informal knowledge*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Jaffe, A. J. & Spirer, H. F. (1987). *Misused statistics: Straight talk for twisted numbers*. New York, NY: Marcel Dekker.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Lajoie, S. P., Jacobs, V. R., & Lavigne, N. C. (1995). Empowering children in the use of statistics. *Journal of Mathematical Behavior, 14(4),* 401-425.

Lajoie, S. P., Lavigne, N. C., & Lawless, J. (1993, April). *The use of hypercard for facilitating assessment: A library of exemplars for reifying statistical concepts*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Lave, J. (1988). *Cognition in practice*. New York: Cambridge University Press.

Lehrer, R. & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14* (1), 69-108.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37* (11), 2098-2109.

Mack, N. K. (1990). Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education, 21* (1), 16-32.

Marascuilo, L. A. & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences*. New York: W. H. Freeman and Company.

National Center for Research in Mathematical Sciences Education & Freudenthal Institute. (Eds.). (in press). *Mathematics in context: A connected curriculum for grades 5-8*. Chicago: Encyclopedia Britannica Educational Corporation.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school.* New York: Cambridge University Press.

Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* New York: Cambridge University Press.

Papert, S. (1980). *Mindstorms.* New York: Basic Books.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences.* New York: Hill and Wang.

Paulos, J. A. (1995). *A mathematician reads the newspaper.* New York: Doubleday.

Resnick, L. B. (1987). Learning in school and out. *Educational Researcher, 16,* 13-20.

Russell, S. J. & Friel, S. N. (1989). Collecting and analyzing real data in the elementary school classroom. In P. R. Trafton & A. P. Shulte (Eds.), *New directions for elementary school mathematics: 1989 Yearbook* (pp. 134-148). Reston, VA: National Council of Teachers of Mathematics.

Saxe, G. B. (1988). Candy selling and math learning. *Educational Researcher, 17,* 14-21.

Schwartz, D. L., Goldman, S. R., & Cognition and Technology Group at Vanderbilt. (in press). *Using anchored instruction to align everyday and mathematical reasoning: The case of sampling assumptions.* In S. P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwartz, D. L., Goldman, S. R., Moore, A. L., Zech, L., Smart, K., Mayfield-Stewart, C., Vye, N. J., & Barron, L. (1994, April). *Adolescent understandings of sampling in the context of a survey.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9-40). Cambridge, MA: Harvard University.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning,* (pp. 465-494). New York: Macmillan.

TERC (in press). *Investigations in number, data, and space* (Series for grades 1-5). Palo Alto, CA: Dale Seymour.